REMARKS / DISCUSSION OF ISSUES

Claims 1-8 and 10-29 are pending in the application wherein claims 9 and 30-32 have been canceled.

Applicants thank the Examiner for acknowledging the claim for priority and receipt of certified copies of all the priority document(s).

The claims in general include certain amendments for one or more non-statutory reasons, such as for better form. Such amendments are not believed to narrow the scope of the claims and no new matter is added.

The Office Action indicates that the information disclosure statement (IDS) filed May 31, 2006 fails to comply with 37 CFR 1.97 and 1.98 for not supplying a copy of an article entitled "Named Faces: Putting Names to Faces" by Houghton. In response a copy of this article by Houghton is enclosed. Accordingly, consideration of WO 00/48395 is respectfully requested.

The Office Action indicates that the full name of Robert Turetsky is not disclosed in the Oat. Upon review of the Declaration on File from PAIR, it appears that full name of Robert Turetsky is not disclosed in the Declaration, a copy of which is enclosed. Clarification is respectfully requested.

The Office Action rejects claims 1-25 under 35 U.S.C. §101 as allegedly directed to non-statutory subject matter. Without agreeing with the Examiner, and in the interest of furthering the prosecution and expediting allowance of the present Application, claims 1 and 26 have been amended for better form that more clearly recites statutory subject matter. It is respectfully requested that the rejection of claims 1-25 under 35 U.S.C. §101 has been overcome and withdrawal of this rejection is respectfully requested.

The Office Action rejects claims 1-5, 8-9, 12, 17-23, 26 and 29 under 35 U.S.C. §102(e) over U.S. Patent No. 6,834,308 (Ikezoye); and rejects claims 6-7, 10-11, 13-21, 24-25 and 27-28 under 35 U.S.C. §103(a) over Ikezoye in view of U.S. Patent No. 6,243,676 (Witteman). It is respectfully submitted that claims 1-8 and 10-29 are patentable over Ikezoye and Witteman for at least the following reasons.

The Office Action alleges on page 6, in rejecting claim 9, that column 4, lines 50-51 of Ikezoye discloses that the extrinsic information source is a film screenplay.

It is respectfully submitted that at best, column 4, lines 50-51 of Ikezoye merely discloses that the extrinsic information source is a film on a DVD.

It is respectfully submitted that Ikezoye, Witteman, and combination thereof, do not teach or suggest the present invention as recited in independent claim 1, and similarly recited in independent claim 26 which, amongst other patentable elements, recites (illustrative emphasis provided):

> a processor configured to correlate the intrinsic data and the extrinsic data for providing a multisource data structure, wherein the audio-visual source is a film and the extrinsic information source is a film screenplay written before production of the film.

These features are nowhere taught or suggested in Ikezoye, Witteman, alone or in combination. Accordingly, it is respectfully submitted that independent claims 1 and 26 allowable. In addition, claims 2-8, 10-25 and 27-29 are allowable at least because they depend from independent claims 1 and 26 as well as for the separately patentable elements contained in each of the dependent claims.

In view of the foregoing, applicants respectfully request that the Examiner withdraw the objection(s) and/or rejection(s) of record, allow all the pending claims, and find the application in condition for allowance. If any points remain in issue that may best be resolved through a personal or telephonic interview, the Examiner is respectfully requested to contact the undersigned at the telephone number listed below.

Enclosure: Article by Houghton
Copy of Declaration from PAIR

Respectfully submitted,

Dicran Halajian
Reg. 39,703
Attorney for Applicant(s)
March 27, 2008

THORNE & HALAJIAN, LLP
Applied Technology Center
111 West Main Street
Phone: (631) 665-5139
Fax: (631) 665-5101

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (includes Reference to PCT International Applications)	ATTORNEY'S DOCKET NUMBER PHNL040163 US

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

the specification of which (check only one item below):

☐ is attached hereto.

☐ was filed as United States application

Serial No

on

and was amended

on

☒ was filed as PCT international application

Number PCT/IB2004/052601

on 30 NOVEMBER 2004

and was amended under PCT Article 19

on (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, § 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:

COUNTRY	APPLICATION NUMBER	DATE OF FILING DAY, MONTH, YEAR	PRIORITY CLAIMED UNDER 35 USC 119
U.S.A.	60/527,476	5 December 2003	YES
Europe	04100622.2	17 February 2004	YES

Combined Declaration For Patent Application and Power of Attorney (Continued) (Includes Reference to PCT International Applications)	Attorneys Docket Number PHNL040163 US

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (List name and registration number)

Jack E. Haken, Reg. No. 26,902 Michael E. Marion, Reg. No. 32, 266 Edward M. Blocker, Reg. No. 30,245	Direct Telephone Calls to: (name and telephone number) (914)332-0222

201	FULL NAME OF INVENTOR	FAMILY NAME **DIMITROVA**	FIRST GIVEN NAME **Nevenka**	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY **Yorktown Heights**	STATE OR FOREIGN COUNTRY **U.S.A.**	COUNTRY OF CITIZENSHIP **Former Yugoslav Republic of Macedonia**
	POST OFFICE ADDRESS	POST OFFICE ADDRESS **3148 Gomer Street**	CITY **Yorktown Heights, CA 10598**	STATE & ZIP CODE/COUNTRY **U.S.A.**
202	FULL NAME OF INVENTOR	FAMILY NAME **TURETSKY**	FIRST GIVEN NAME **Robert**	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY **Passaic**	STATE OR FOREIGN COUNTRY **U.S.A.**	COUNTRY OF CITIZENSHIP **U.S.A.**
	POST OFFICE ADDRESS	POST OFFICE ADDRESS **84 Crescent Ave.**	CITY **Passaic, NJ 07055**	STATE & ZIP CODE/COUNTRY **U.S.A.**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 if Title 18 of the United states Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201	SIGNATURE OF INVENTOR 202
x [signature]	
DATE x 12/9/2004	DATE

U.S. DEPARTMENT OF COMMERCE- Patent and Trademarks Office

(July 1994)

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (Includes Reference to PCT International Applications)	ATTORNEY'S DOCKET NUMBER PHNL040163 US

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

the specification of which (check only one item below):

☐ is attached hereto.

☐ was filed as United States application

Serial No

on

and was amended

on

☒ was filed as PCT international application

Number PCT/IB2004/052601

on 30 NOVEMBER 2004

and was amended under PCT Article 19

on (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, § 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:

COUNTRY	APPLICATION NUMBER	DATE OF FILING DAY, MONTH, YEAR	PRIORITY CLAIMED UNDER 35 USC 119
U.S.A.	60/527,476	5 December 2003	YES
Europe	04100622.2	17 February 2004	YES

Combined Declaration For Patent Application and Power of Attorney (Continued)
(Includes Reference to PCT International Applications)

Attorneys Docket Number
PHNL040163 US

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (List name and registration number)

Jack E. Haken, Reg. No. 26,902
Michael E. Marion, Reg. No. 32, 266
Edward M. Blocker, Reg. No. 30,245

Direct Telephone Calls to:
(name and telephone number)
(914)332-0222

201	FULL NAME OF INVENTOR	FAMILY NAME DIMITROVA	FIRST GIVEN NAME Nevenka	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY Yorktown Heights	STATE OR FOREIGN COUNTRY U.S.A.	COUNTRY OF CITIZENSHIP Former Yugoslav Republic of Macedonia
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 3148 Gomer Street	CITY Yorktown Heights, CA 10598	STATE & ZIP CODE/COUNTRY U.S.A.
202	FULL NAME OF INVENTOR	FAMILY NAME TURETSKY	FIRST GIVEN NAME Robert	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY Passaic	STATE OR FOREIGN COUNTRY U.S.A.	COUNTRY OF CITIZENSHIP U.S.A.
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 84 Crescent Ave.	CITY Passaic, NJ 07055	STATE & ZIP CODE/COUNTRY U.S.A.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 if Title 18 of the United states Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201	SIGNATURE OF INVENTOR 202
DATE	DATE 1/20/05

Named Faces: Putting Names to Faces

Ricky Houghton, Carnegie Mellon University

THE NAMED FACES SYSTEM AUTOMATICALLY ACCUMULATES A DATABASE OF PAIRED NAMES AND FACES. WHEN USERS SUBMIT IMAGES CONTAINING A FACE, THE SYSTEM ATTEMPTS TO RETURN THE NAME FOR THAT FACE.

FOR MULTIMEDIA SEARCH AND retrieval systems, automatically labeling people in video has many benefits, including support for multimodal queries and story summarization. In the first case, a simple text query is sometimes inadequate for searching multimedia databases. For example, users wishing to find all instances of President Clinton discussing taxes might be overwhelmed with results that involve taxes and Clinton, but not necessarily Clinton discussing taxes. If a video database has named the people appearing in the footage, more refined queries or multimodal queries are possible. For example, `search: face = "President Clinton" text = taxes` limits the search to news stories that contain President Clinton's face.

For story summarization, a system's ability to automatically identify the persons appearing in video will increase its ability to summarize a story. A system should be able to take advantage of the superimposed text, which provides much information that is not in the audio data or speech. Many times a video clip references a person by his or her title—for example, "White House spokesperson" instead of "Mike McCurry." In these cases, being able to extract the name of the person shown might be an important factor for summarization. For instance, sometimes video participants clearly define the topics. Extracting the names of these participants becomes especially important if they are not mentioned in the news story but are seen in the video. In analyzing six hours of CNN newscasts, I found that 39% of the faces were labeled—17% in audio and 22% in superimposed text. An accurate system should be able to correctly label more than 39% of the faces in a news broadcast.

To provide automatic labeling of faces in video, I've developed Named Faces, a fully functional automated system that builds a large database of name-face association pairs from broadcast news. This article describes how my system detects and recognizes superimposed text in the video, then verifies or repairs the text by comparing it with a large list of automatically generated names found in news stories. Faces found in the video where superimposed names were recognized are tracked, extracted, and associated with the superimposed text. With Named Faces, users can submit queries to find names for faces in video images.

Creating a database of names and faces

The creation process consists of

1. retrieving and analyzing data from the Web,
2. creating a list of names,
3. video optical character recognition (VOCR),
4. dynamic programming to improve OCR results, and
5. extracting the best face from the video and inserting it in the Named Faces Database.

Figure 1 diagrams this process.

1094-7167/99$10.00 © 1999 IEEE

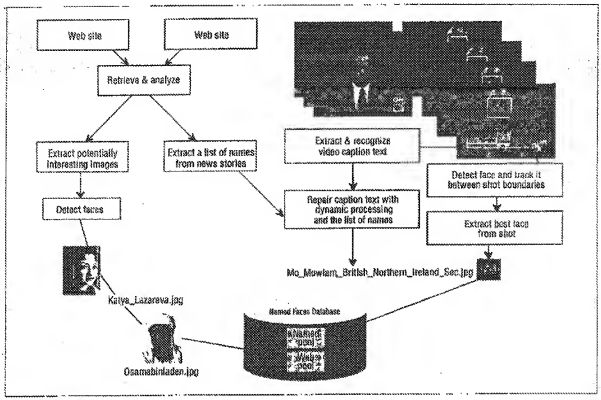

Figure 1. The Named Faces creation process.

Retrieving and analyzing Web data. Major news sources have a significant presence on the Web and provide information that is useful for building the Named Faces Database. Named Faces mines CNN, ABC, and Associated Press Web pages during the early morning hours for potentially useful images of people and names of people featured in news stories. After Named Faces downloads each page, it parses the HTML data first for the images and then for the names.

Finding the images. Potentially useful images have filenames extracted from the HTML page that appear descriptive. Nonuseful images have filenames that contain only numbers or two alpha characters. Figure 2 contains samples of extracted HTML lines.

Although this method mines useful data, some seemingly useful tags are confusing or not useful. For instance, Princess Diana's bodyguard appears in the database as "Diana Jones." This clearly is not his name; instead, it refers to Jones' being related to Diana in a news sense. Similarly another image was labeled "Microsoft," while the image contains Bill Gates' face. Figure 3 contains sample images retrieved from the Web.

Once all the images have been retrieved, the system passes them to Facelt, a commercial face detector (http://www.facelt.com), to see if the image contains a face. Images for which a face and a descriptive name exist are then sent to the Named Faces Database. Images extracted from the Web are maintained in the *Web pool*, a separate division of the database, for tracking and future pruning.

Finding the names. To find the names of people in news stories, Named Faces passes the retrieved HTML pages to a Perl script that uses several widely available HTML parsing modules. The output of the modules is text from the page, clean of any HTML tags. However, this text is not necessarily free of non-news text; titles or long advertisements might remain. Each site has its own formatting rules, which Named Faces exploits to remove as much non-news text as possible. This final filtering step is not required, but it makes the data cleaner. The final output is stored on disk for later processing.

Creating the name list. Named Faces uses the filtered news stories as a source for generating a list of names. It uses two methods to build this list. The first is the direct extraction of names from HTML. This method is possible for one of the news Web sites, which provides transcripts for a good portion of its

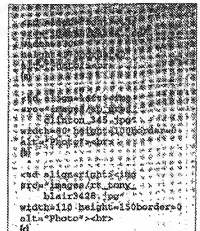

Figure 2. Sample lines from HTML data retrieved from Web sites. Some names require removal of nonalpha characters. ap_ means the data is from the Associated Press, and xt_ means it is from Reuters. Named Faces removes these tags and all numbers. Names extracted for these images after filtering: (a) Gates; (b) Prez Clinton; (c) Terry Moran.

Figure 3. Sample images retrieved automatically from the Web. Tags associated with these images are (from left to right, top to bottom) Pres Clinton, Lewinsky, Clinton Hillary, Yelstin Smile, Diana, Chernomyrdin.

broadcasts. Although these transcripts are no more important than any other news source, their format is more useful than that of normal news stories. Each transcript contains extra information that simplifies name extraction. Figure 4 contains sample filtered transcript data. With this method, Named Faces has gathered 3,600 names. This includes seemingly duplicated data such as William Jefferson Clinton, which is different from William J. Clinton. A few people have multiple names; each of these instances is counted as unique.

The second method exploits online US Census data (http://www.census.gov). Named Faces uses three lists from the Census Bureau: last names, female first names, and male first names. The names are ordered by their frequency. The system employs the brute-force approach of comparing each possible name combination from the Census Bureau data to the extracted news stories. It drops the bottom 1% of names from each set of census data because they are infrequent and they produce a lot of false positives.

The combination of the two extraction processes has produced 17,000 unique names from more than 24,000 news stories. That is, 17,000 unique names occurred in stories that are part of the archived news stories gathered from the Web. Named Faces ignores possible name combinations from census data that are not found in news archives.

I am pursuing a third method: using a named-entity tagger—in particular, Mitre's Alembic text-processing system.[1] This approach would increase the number of names extracted.

Video optical character recognition. Text overlaid on the screen, sometimes referred to as *Chyron text* or *caption text*, is very useful for generating a set of name-image pairs. A broadcaster adds Chyron text to provide names, titles, topics, and location information to the viewer. Because the text appears in the video as part of the image and not part of a data stream such as closed-caption data, Named Faces uses the Informedia VOCR system[2] to extract this information from videos of news broadcasts.

Given the number of frames in typical broadcast news, processing every video frame for text is not computationally feasible. So, the VOCR system first performs a quick text-region detection. Using horizontal differential filtering, it searches for horizontal rectangular structures of clustered sharp edges. Such structures usually indicate text. The system then sequentially filters these potential text regions across all detected frames, effectively increasing each caption's resolution. Further filtering reduces background noise. The system then extracts the potential text region as a TIFF image and submits the image to TextBridge (http://www.scansoft.com), a commercial OCR program, for the final stage of text recognition.

Repairing VOCR results. Unfortunately, state-of-the-art VOCR techniques have a word error rate of 65%, which is not accurate enough for reliable name-face association. With such an error rate, the probability of correctly recognizing a first-name, last-name pair drops to approximately 12%.

To repair the VOCR output, Named Faces uses *dynamic programming* (also called *dynamic time warping*) with the list of names. DP consists mainly of a dynamic comparison between a test and a reference pattern.[3] It is very useful for computing the similarity (that is, a fuzzy match) of two strings of characters. In Named Faces, this method compares each line of VOCR output against each name extracted from the news story archive.

Instead of performing a DP match against each word, Named Faces performs character-level matching. To achieve this, it substitutes each group of spaces with a single "_" character. Then, it removes all remaining spaces. Finally, it inserts a space between characters. By performing these steps, Named Faces treats each character as a word. DP match-

Figure 4. Filtered transcript data. Capitalization comes from the source; the bold print has been added to highlight the formatting.

Table 1. Some visual optical character recognition output and the results after repairing the output using dynamic programming and a list of named entities.

VOCR output	Repaired output
A Time Warner Compan– Join Chen	JOE CHEN
Behind Closed Door– Paul McNulty Counsel to Judiciary Cmte	PAUL MCNULTY, COUNSEL TO JUDICIARY CMTE.
Rep Ste phep Bu–r H Indiana FT Hyds and Seek	REP. STEPHEN BUYER (R), INDIANA
Rep Tom DeL–y R Majority Wh. ip Joe Lockhar W–H Deputy Press Secretary	REP. TOM DELAY (R), MAJORITY WHIP JOE LOCKHART, W.H. DEPUTY PRESS SECRETARY
Roger Clinton President Clinton S Brother Rep Este ban Torres D California	REP. ESTEBAN TORRES (D), CALIFORNIA
Candy Cro vi ~ fil fj–FF~~ V Clintons Shadow	
Sen_ Paul We us tone D Minnesota	SEN. PAUL WELLSTONE (D), MINNESOTA
Prof_ Ste phoP Sa–lzburg George W–S hington Univ Law Schno	PROF. STEPHEN SALTZBURG, GEORGE WASHINGTON UNIVERSITY LAW SCHOOL
Melanie Kibler Congressional Staffe	MELANIE KIBLER, CONGRESSIONAL STAFFER

ing returns Match, Insertion, or Deletion for each character in the data set (the VOCR output) relative to the string it is compared against, and returns a total score based on the distance between the two strings. Match indicates that a letter in the VOCR output string matches a letter in the other string. Insertion means that a letter in the VOCR output string is not in the other string. Deletion means that a letter in the other string is not in the VOCR output string. The DP match's output also labels where the matches, insertions, and deletions occur. This is useful for a finer level of thresholding; Named Faces currently thresholds the match's score with insertions and deletions penalized equally. Table 1 shows some VOCR output with the corrected output after the DP match.

To test DP-match repairing, I selected two 60-minute CNN news broadcasts, which produced 177 lines of VOCR results. Of these lines, 32 represented the name of the person on the screen, excluding commercials. DP-match repairing decreased the error rate to 45%, from the standard VOCR error of 65%. That is, by comparing the VOCR output with the list of names generated earlier, Named Faces can repair VOCR results when they closely match a name in the name list. I define closeness as a low percentage of characters in the results that do not match those in the name.

A better application of the DP-match program could probably produce significant gains. For example, a cleverer DP match would weight each character based on its "confusability" with potential matches. For instance, "I" and "t" are highly confusable in the current VOCR system, so the DP match should not penalize a match that has misrecognized an "I" as a "t." It should however, heavily penalize an "m" or some other character that is not part of the highly confusable set. This step, which I am incorporating into Named Faces, should improve the DP match's performance.

Extracting and submitting faces. Once Named Faces has repaired the VOCR results, it uses the resulting data to extract relevant faces from the video. The VOCR output includes start and end frames for when the overlaid text appeared in the video. A name found in the output will likely belong to the face in these frames. Bounded by these frames, the system tracks the face every 15th video frame (or every half second) to find the highest-scoring face. It then associates this face with the name found in the VOCR output.

I've improved basic face extraction by using *shot breaks*. Standard Informedia processing uses comparative difference measures[4] to segment the video into groups of similar images, called *shots*.[5] Images with small color histogram disparity are considered relatively equivalent. The point where two shots meet is a *shot break*.

By using this extra information, Named Faces can search not just the frames bounded by the VOCR data but also the entire shot in which the face is likely to have appeared. This is helpful because by definition a shot contains roughly the same image for the entire range. Given that the VOCR output typically appears on the screen for less than 50% of the shot range, the wider boundaries of shot breaks are quite useful. Tracking is now performed over the entire range of the shot boundaries. Named Faces selects the face that receives the highest confidence from Facelt and extracts it from the group.

Named Faces gives the extracted images the name of the corrected VOCR output and inserts this data into the *named pool* in the Named Faces Database.

Querying the database

Once a database is built, users can find names for unlabeled images by submitting the images to the Named Faces Database. Images are sent to the database through a simple socket interface. Facelt compares a new image with each image in the database. It returns an *N*-best list of names, with a score representing the quality of the match to each name. Figure 5 diagrams the querying process.

The combination of the Web and named pools permits searching just Web data, just VOCR data, or both. For this article's results, the search included both sets of data, which is the system's default.

Submitting multiple images. I've extended this model to improve the labeling's accuracy. Because a face might have different views or expressions in a shot, the image that is chosen for recognition might negatively affect the results. Submitting multiple facial images instead of one increases the recognition's reliability. For example, a face turning from right to left in a shot might initially be matched with the wrong person. However, other sampled faces from this shot will probably not all be matched with that person. Submitting multiple images should reduce false alarms.

For a submission of multiple images, Named Faces returns an *N*-best list of scores for each image. These lists must be merged or compared before a result can be determined. For combining multiple *N*-best lists, Named Faces uses the *weighting models rank* method, proposed by K. Markov and Seiichi Nakagawa.[6] This method weights each name in the *N*-best lists with a function based on the name's rank in each list. Each name's score is the sum of its weighted scores across the full shot. For a shot with T faces, the score of each name, S_n, is the sum of the sums of rank weighted scores, $w(N - n) * S_{i-n}$, where N is the size of the *N*-best list. If the name with the highest score, $\max(S_i)$, exceeds the matching threshold, Named Faces returns that name *Oname(i)* as the match. The equation for combining the multiple *N*-best lists is

$$S_i = \sum_{t=1}^{T} \sum_{n=1}^{N} w(N-n) s_{i(t)n} .$$

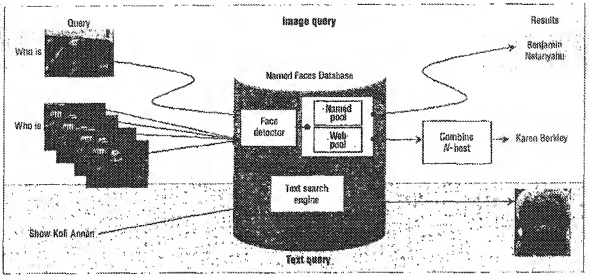

Figure 5. The query process for images or text.

Testing and results. I tested three query methods: single-frame query, multiple-image query with linear weighting, and multiple-image query with exponential weighting. To test these methods, I submitted images from two news broadcasts to the Named Faces Database. I then compared the results with those of a human that performed the same task.

Given the closed-caption data and the overlaid caption text, the human correctly labeled 87% of all the faces in the video, with no false alarms. (The human was unable to label the other 13%.) Single-frame query correctly labeled 53% of the faces, with a 15% false-alarm rate. The *N*-best methods performed roughly equally. The linear *N*-best method correctly labeled 41%, with 2% error. The exponential *N*-best method correctly labeled 42%, with 3% error.

I also took into account the database's breadth by measuring the number of different faces labeled compared to the total number of different faces. (By different, I mean the faces of different people as opposed to multiple images of the same person.) The single-frame method correctly labeled at least one face of 61% of the different people. That is, it did not label 39% of the people in the video at all. Both *N*-best methods correctly labeled at least one face of 53% of the different people. The human correctly labeled at least one face of 100% of the different people.

NAMED FACES GOES BEYOND simple face detection and recognition to content extraction or description. It can become more accurate over time with unsupervised TV monitoring. It is reliable enough to be integrated into an intelligent multimedia database retrieval system such as Informedia.[7]

I have also tested Named Faces on several different styles of broadcast news, including foreign broadcasts recorded from Scola, a nonprofit organization that broadcasts international programs by satellite around the world (http://www.scola.org). At the time of the American bombing of Sudan (August 1998), Named Faces processed Croatian, Mexican, and Spanish news broadcasts. In each of the three broadcasts, the system correctly labeled images of President Clinton, while other faces that were not in the database did not return false positives. I believe that automatic labeling of people occurring in video can be useful for several aspects of large multimedia database systems, such as summarization, multimodal queries, and six-degrees search.

One near-term goal is to incorporate Name-It's co-occurrence processing (see the sidebar). Given that Named Faces now processes and handles names as entities instead of proper nouns, the co-occurrence method should be especially useful when video caption text is not available.

Another near-term goal is to increase the database's size. The current database contains approximately 900 named images that were gathered from 50 of Informedia's 1,000-plus hours of online news footage. A significant increase in the database's size should improve results dramatically.

One of the more interesting extensions to

Related work

Name-It, a system developed at Carnegie Mellon University, performs a similar association using *video optical character recognition* techniques to recognize the superimposed caption information displayed on the screen, as well as using closed-caption (teletype) data broadcast with the video.[1] With only these two sources of information, associating names and faces is difficult.

Because VOCR techniques have a word error rate of 65%, extracting names reliably is very difficult without some form of regular Name-It cannot proper nouns from the closed-caption data and uses a co-occurrence algorithm to determine whether similar faces co-occur with an extracted noun. However, Name-It has problems extracting proper nouns as named entities. For instance, instead of extracting "Bill Clinton," the system can only extract "Clinton" because Bill might be a verb. Furthermore, the system does not know about compound names as a unit; it deals with proper nouns as single words. I designed Named Faces (see the main article) largely to address the Name-It system's shortcomings mentioned here.

References

1. S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and Detecting Faces in Video by the Integration of Image and Natural Language Processing," *Proc. Int'l Joint Conf. Artificial Intelligence*, Morgan Kaufmann, San Francisco, 1997.

How to Reach Us

Writers

For detailed information on submitting articles, write for our Editorial Guidelines (m.davis@computer.org), or access http://computer.org/intelligent/edguide.htm.

Letters to the Editor

Send letters to

Managing Editor
IEEE Intelligent Systems
10662 Los Vaqueros Circle
Los Alamitos, CA 90720
dprice@computer.org

Please provide an e-mail address or daytime phone number with your letter.

On the Web

Access http://computer.org/intelligent/ for information about *IEEE Intelligent Systems*.

Subscription Change of Address

Send change-of-address requests for magazine subscriptions to address.change@ieee.org. Be sure to specify *Intelligent Systems*.

Membership Change of Address

Send change-of-address requests for the membership directory to directory.updates@computer.org.

Missing or Damaged Copies

If you are missing an issue or you received a damaged copy, contact membership@computer.org.

Reprints of Articles

For price information or to order reprints, send e-mail to m.davis@computer.org or fax (714) 821-4010.

Reprint Permission

To obtain permission to reprint an article, contact William Hagen, IEEE Copyrights and Trademarks Manager, at whagen@ieee.org.

the Named Faces Database is combining audio data from the video stream to improve the system. The audio portion of the video has a high probability of belonging to the person appearing on the screen. Determining when this is the case and accurately labeling the audio as belonging to that person is feasible. I am extracting audio data from regions where Named Faces has labeled a face. I've captured a small database of more than 100 labeled regions of audio. I'll further process this database to develop a Named Voice system that automatically learns the voices of people featured in the news, similar to Named Faces. The second step will be to combine the audio and video data to allow queries consisting of audio and video. Combining these data should increase the system's accuracy. Pruning should become more accurate because facial images that are confusable will probably not have voices that are equally confusable. Daniele Falavigna and Roberto Brunelli have demonstrated significant gains by using audio and visual features together.[8]

Another possible use of the audio data is to label the face image as either Male or Female/Child (distinguishing female and children's voices is difficult). I plan to accomplished this by extracting features from the audio track that belongs with each image. This will provide one more piece of information that might be useful for summarization or retrieval.[3]

Finally, I'll incorporate a text search engine into the Named Faces Database so that a text query can be made against the names of the faces in the database. ✉

Acknowledgments

This article is based on work supported by the National Science Foundation, DARPA, and NASA under NSF Cooperative Agreement IRI-9411299. I thank the Informedia Project partners for the video data and other support, and the Informedia Project team members for all the underlying work in image processing, natural-language processing, speech recognition, library data creation, library delivery, and project management. The partners and team members are listed at http://www.informedia.cs.cmu.edu/ along with information on the Informedia Project.

References

1. D. Day, J. Aberdeen, and L. Hirschman, "Mixed-Initiative Development of Language Processing Systems," *Proc. Fifth Conf. Applied Natural Language Processing*, Assoc. for Computational Linguistics, Washington D.C., 1997.

2. T. Sato et al., "Video OCR for Digital News Archive," *Proc. CAIVD '98: 1998 Int'l Workshop on Content-Based Access of Image and Video Databases*, IEEE Computer Soc. Press, Los Alamitos, Calif., 1998, pp. 52–60.

3. H. Sakoe and A. Chiba, "Dynamic Programming Approach to Continuous Speech Recognition," *Proc. Int'l Congress on Acoustics*, Paper 20-C-13, 1971.

4. H. Zhang, A. Kankanhalli, and S. Smoliar, "Automatic Partitioning of Full-Motion Video," *Multimedia Systems*, Vol. 1, No. 1, Jan. 1993, pp. 10–28.

5. R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video Abstracting," *Comm. ACM*, Vol. 40, No. 12, Dec. 1997, pp. 55–63.

6. K. Markov and S. Nakagawa, "Frame Level Likelihood Normalization for Text-Independent Speaker Identification Using Gaussian Mixture Models," *Proc. Int'l Conf. Speech and Language Processing*, IEEE Press, Piscataway, N.J., 1996, pp. 1764–1767.

7. H.D. Wactlar et al., "Lessons Learned from Building a Terabyte Digital Video Library," *Computer*, Vol. 32, No. 2, Feb. 1999, pp. 66–73.

8. D. Falavigna and R. Brunelli, "Person Recognition Using Acoustic and Visual Cues," *Proc. ESCA Workshop on Automatic Speaker Recognition Identification Verification*, European Speech Communication Assoc., Bonn, Germany, 1994, pp. 71–74.

Ricky Houghton is the manager of process engineering for the Informedia project at Carnegie Mellon University. His technical interests are speech recognition, speaker identification, face detection and recognition, and their application to human-agent or human-robot interfaces. He previously developed a speech-recognition-based speech-training aid for hearing-impaired users at Indiana University. He has also worked on face detection and tracking with active cameras and has developed speaker-identification software for a major Japanese computer company. He received his BS in computer science from Indiana University. He is a member of the IEEE Computer Society. Contact him at the School of Computer Science, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213; ricky.houghton@cs.cmu.edu.